Integrated User Products and Services Support: Virtual Domain Application Data Center (VDADC) and Correlative Studies on Global Data Sets

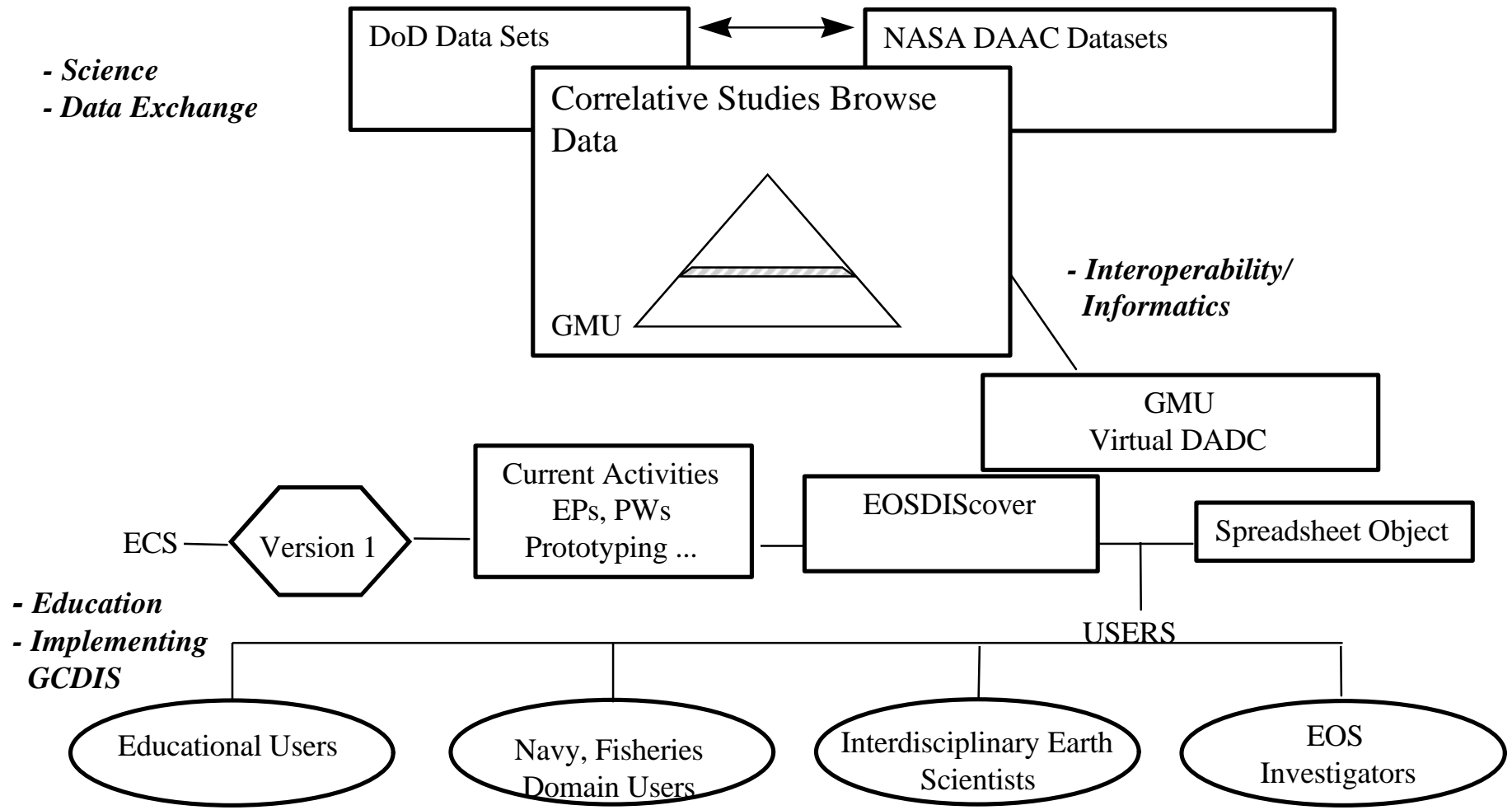Menas Kafatos, Duane King

Ruixin Yang, Hank Wolf

Center for Earth Observing and Space Research (CEOSR)

Institute for Computational Sciences and Informatics

George Mason University

VISION
INTEGRATED USER PRODUCTS AND SERVICES SUPPORT

DoD Data Sets ⟷ NASA DAAC Datasets

Correlative Studies Browse Data

GMU

- *Science*
- *Data Exchange*

- *Interoperability/ Informatics*

GMU
Virtual DADC

ECS — Version 1 — Current Activities EPs, PWs Prototyping ... — EOSDIScover — Spreadsheet Object

- *Education*
- *Implementing GCDIS*

USERS

Educational Users

Navy, Fisheries Domain Users

Interdisciplinary Earth Scientists

EOS Investigators

I. Integrated Support

1. Science

Based on common interests, we focus our efforts on global data sets and associated interdisciplinary Earth science. We plan to use existing DAAC data sets and bring in other (e.g. DoD) in-situ and remote sensing data sets.

2. Data exchanges

Data exchanges will focus on MTPE data sets and other correlative (e.g. DoD) data holdings.

The next three areas are all part of GMU's strategic focus on assisting NASA to expand the user base of EOS and were a strong part of the 1994 IAS recommendations including the DADC concept. Also, based on the August 1995 NRC study and overall emphasis at NASA to expand EOSDIS usage, it is likely that efforts to expand the EOSDIS user communities and combine EOSDIS with GCDIS will increasingly become important.
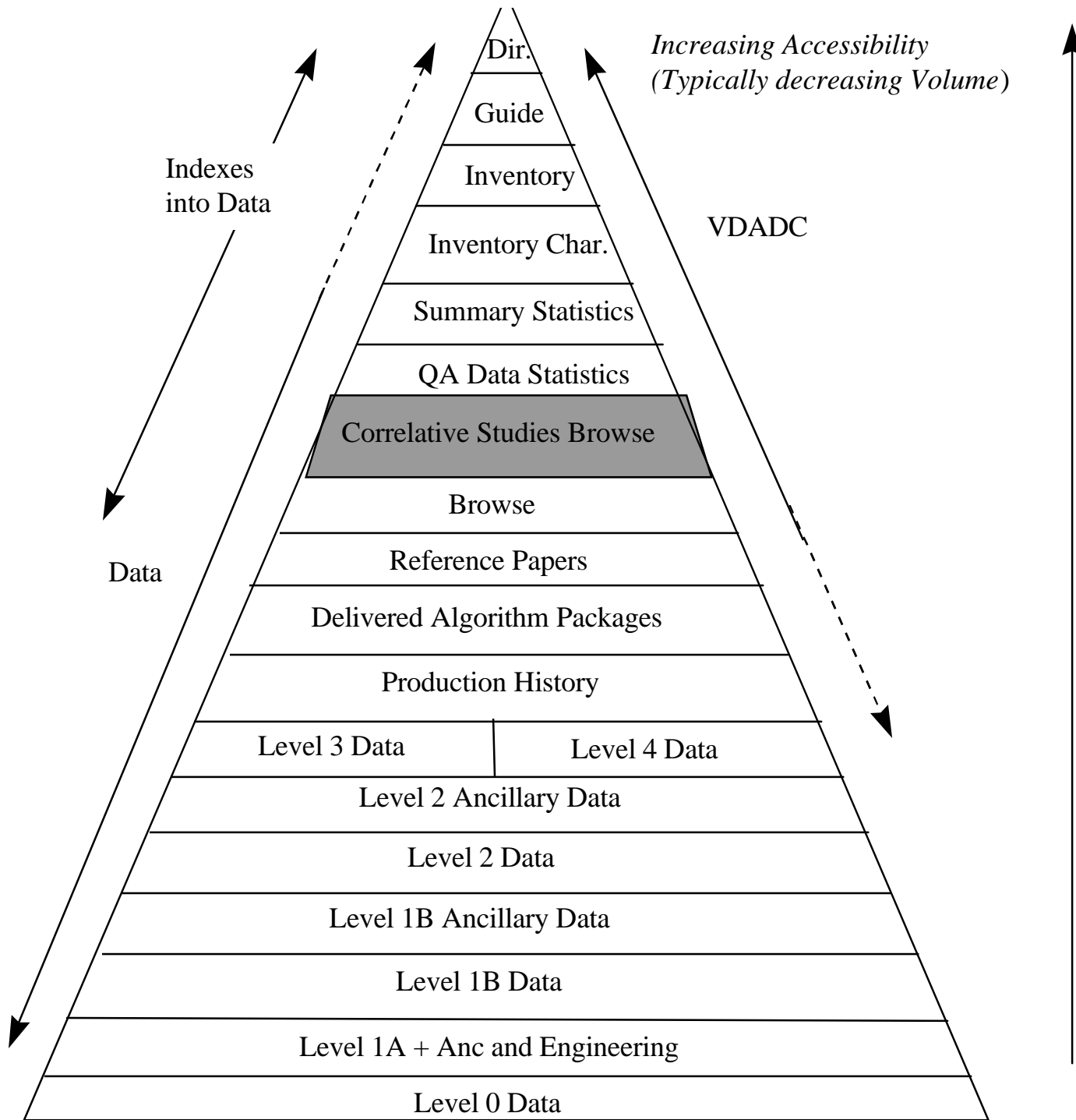
3. Interoperability/Informatics

This is related to exchanging specific data products and creating a federated system of diverse users -- it is crucial for realizing GCDIS as well. We will initially concentrate on correlative studies browse data -- these would likely be of greater interest to domain users such as fisheries and educational users (see our reivsed "Hughes Pyramid".

4. Education

Here we will concentrate on producing correlative studies browse data for the educational communities. We have global data sets from the GSFC DAAC (see below). We could also collaborate with others for producing educational products from lower level, larger volume scientific data.

5. GCDIS/ESIS

This was an important result of the 1994 IAS -- combining ESODIS with GCDIS. A federation of a DoD center (e.g. NRL) with a NASA center (GSFC DAAC) with a university (GMU) and a corporation (HITC) could provide implementation of GCDIS in the Wasington, DC metropolitan area. This would be in line with the proposed NRC Earth Science Information Systgem (ESIS).

Dir.

Guide

Inventory

Inventory Char.

Summary Statistics

QA Data Statistics

Correlative Studies Browse

Browse

Reference Papers

Delivered Algorithm Packages

Production History

Level 3 Data | Level 4 Data

Level 2 Ancillary Data

Level 2 Data

Level 1B Ancillary Data

Level 1B Data

Level 1A + Anc and Engineering

Level 0 Data

Indexes into Data

Data

Increasing Accessibility
(Typically decreasing Volume)

VDADC

II. Implementation of a
Virtual Domain Application
Data Center

COMBINING THE DATA-CENTRIC and USER-CENTRIC VIEWS in the EOSDIS Program

Duane King, Menas Kafatos
Center for Earth Observing & Space Research
Institute for Computational Sciences & Informatics
George Mason University

DOMAIN APPLICATION DATA/DISTRIBUTION CENTER (DADC)

THE DADC WILL SERVE

USER COMMUNITIES

AS A FRONT END

TO EOSDIS

VIRTUAL DADC (VDADC) CONCEPT

Implement DADC on the network

Allow user-specific "pull" demand implementation

Use the VDADC as a front end to EOSDIS that seamlessly has direct access to EOSDIS datasets and products

Store data set pointers

Store multiple pointers per data set to increase reliability

Store selected data products: e.g. correlative browse data binned to uniform specifications (1• x 1• , monthly averaged, etc.)

Some data products will be generated only when requested

Create a roadmap to where the data can be found: local cache (if available), DADCs, ADCs, SCFs, (issue of network vs. local cache)

Maintain local subsets in cache memory of the most frequently used data sets and/or newly created data sets (e.g. some Level 2, Level 3, Level 4 data sets)

DESIGN AND OPERATION CONCEPT

Data taxonomy would be based upon logical groupings such as sea surface temperature (SST), NDVI, or moisture

User request specifies the search parameters based on the designed outcomes, i.e. a data object based search: data type - NDVI and set of parameters - NDVI > 0.5, location, time, etc.

User COTS or other readily available softwares plus VDADC front end - e.g. Netscape and search engine in World Wide Web environment

Data map pointers specify data set details (metadata) and precise data set location(s)

Data storage tracks recent usage and related statistics and a file locking mechanism to prevent other requestors from retrieving incomplete dataset and manage local cache

Limited local processing

Responsible for storing and maintaining domain-specific maps to data and selected data products

DESIGN AND OPERATION CONCEPT
(Continued)

Data retrieval uses anonymous FTP daemon program (FTPD) and determines if retrieved dataset should be stored

File conversions performed on the fly based on file's post-fix or lack of one, e.g. data set name/file name

Data search engine consists of WWW fill-in forms that aid the user in formulating their query and preparing for elementary data analysis, e.g. browse images

User's Web browser is set up to spawn an external viewer, i.e., one that reads HDF, to perform additional processing on the returned data set

Data can be returned as one or more files based upon file size and could also be prepared for further analysis at a data center

VDADC IMPLEMENTATION

1. Data Search Engine: The data search mechanism will consist of a number of WWW fill in forms. These forms will query the user for the type and parameters of the data request.

2. Mapping Information: The data stored at the VDADC is a set of pointers to where data can actually be found. Each file would hold the true location of a data set and provide metadata (directory, guide, inventory and other pertinent data) or browse data information. The format of these files would be free format ASCII strings.

3. Data Storage: The VDADC would store the pointers to all the actual data and would also store a certain amount of "real" data (e.g. subsets of L3 and L4 as well as browse data binned to uniform specifications). Since requests for data will be generated asynchronously, it will be necessary to utilize a file locking mechanism which will indicate the validity of the data being stored.

4. Data Mapping: The data mapping is an ASCII file consisting of keyword = value pairs. The data is then downloaded based on the name and platform as well as the location.

5. Data Retrieval: The mechanism for data retrieval will be anonymous ftp. The ftp daemon program (FTPD) will be responsible for following data set pointers and obtaining and returning the correct data set to the user.

DATA EVALUATION

Use Public Domain Web Interface Applications

VALUE ADDED

User builds a space that contains:

Links to data

Filters to be applied per data item (e.g. time constraints, unit conversion, etc.)

Statistical and visualization packages

Massaged data items to form animations
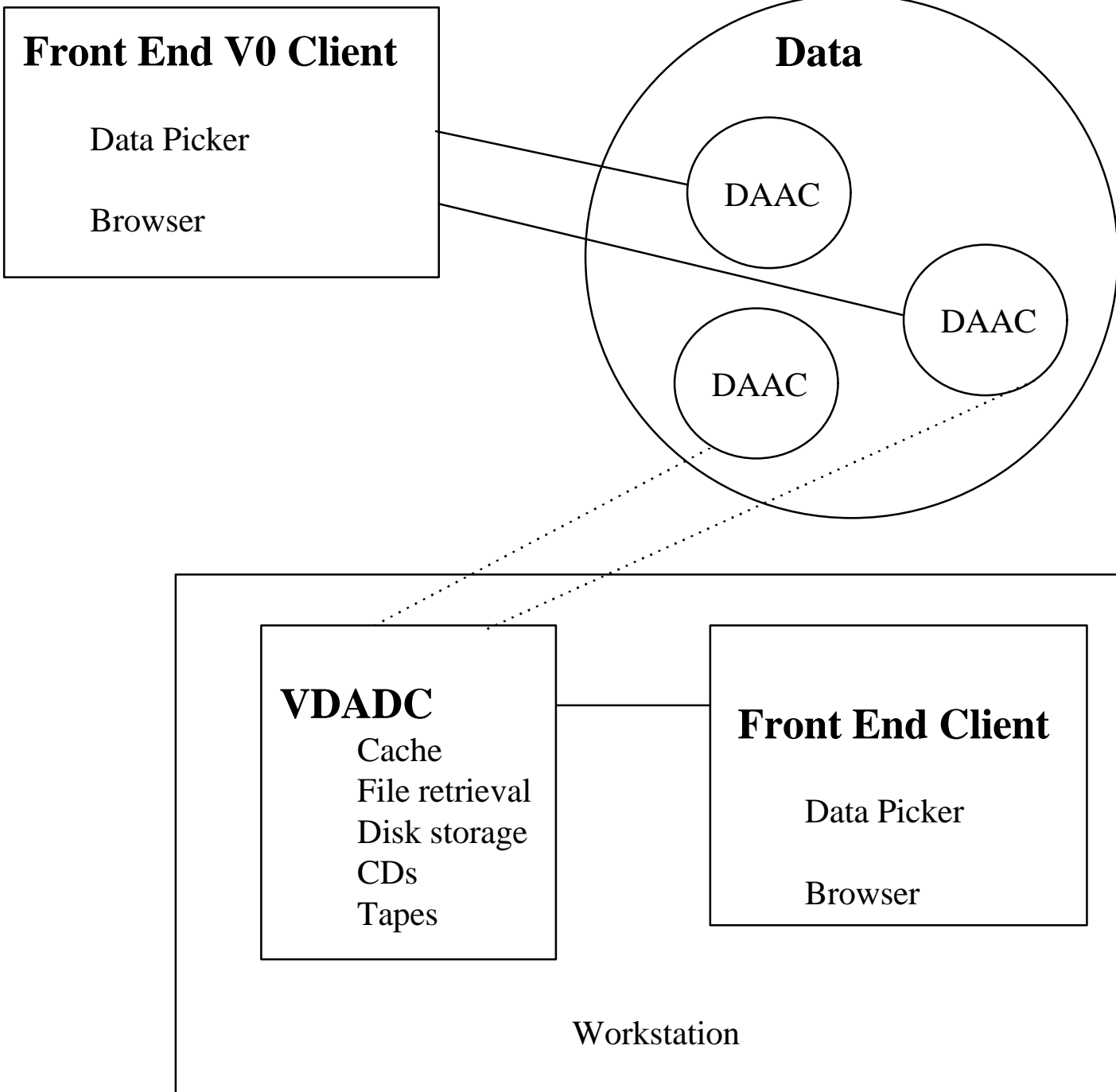
Filters to process the animations

VDADC

Provides for the logical creation of a DAAC with a minimal amount of disc space
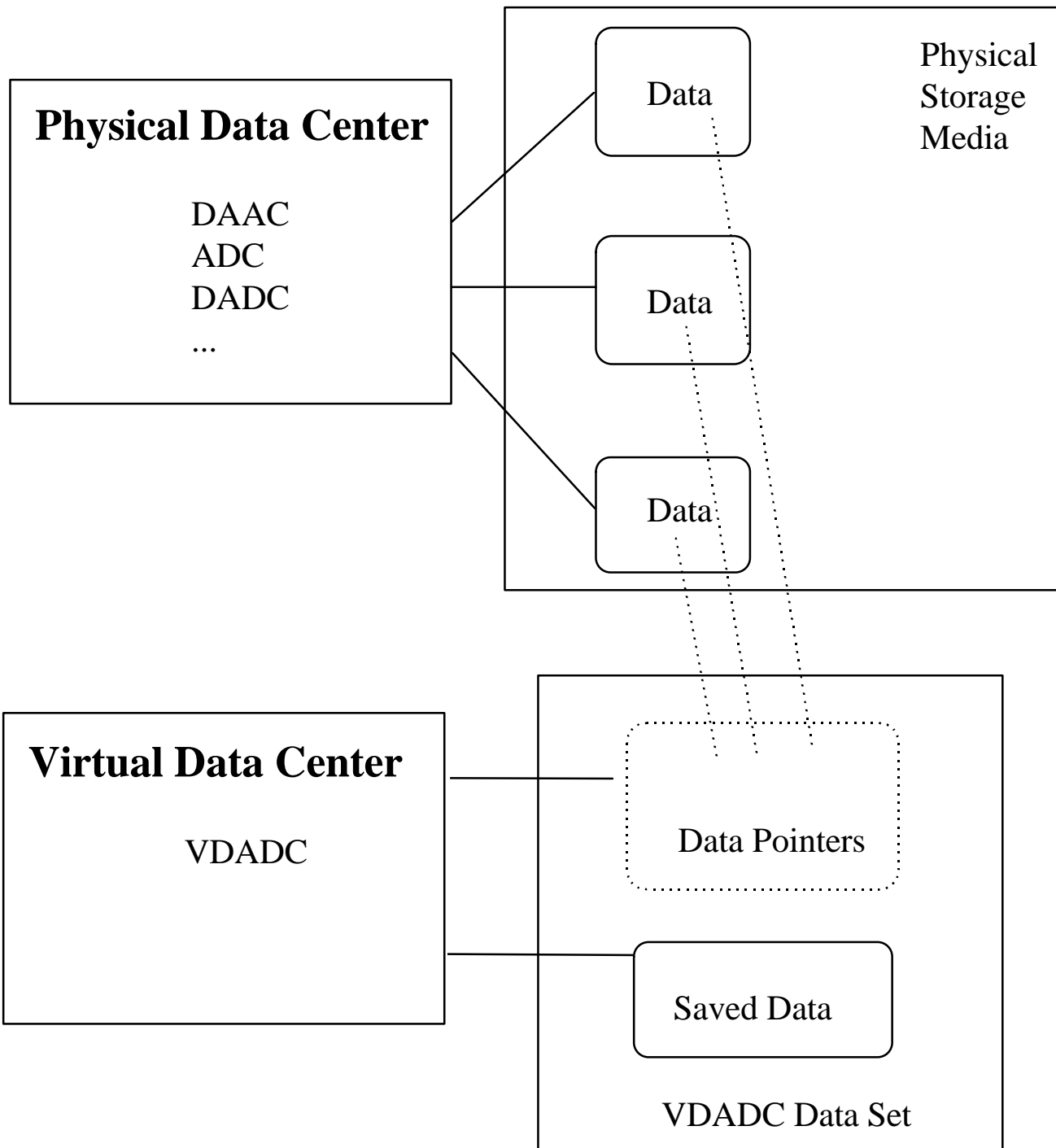
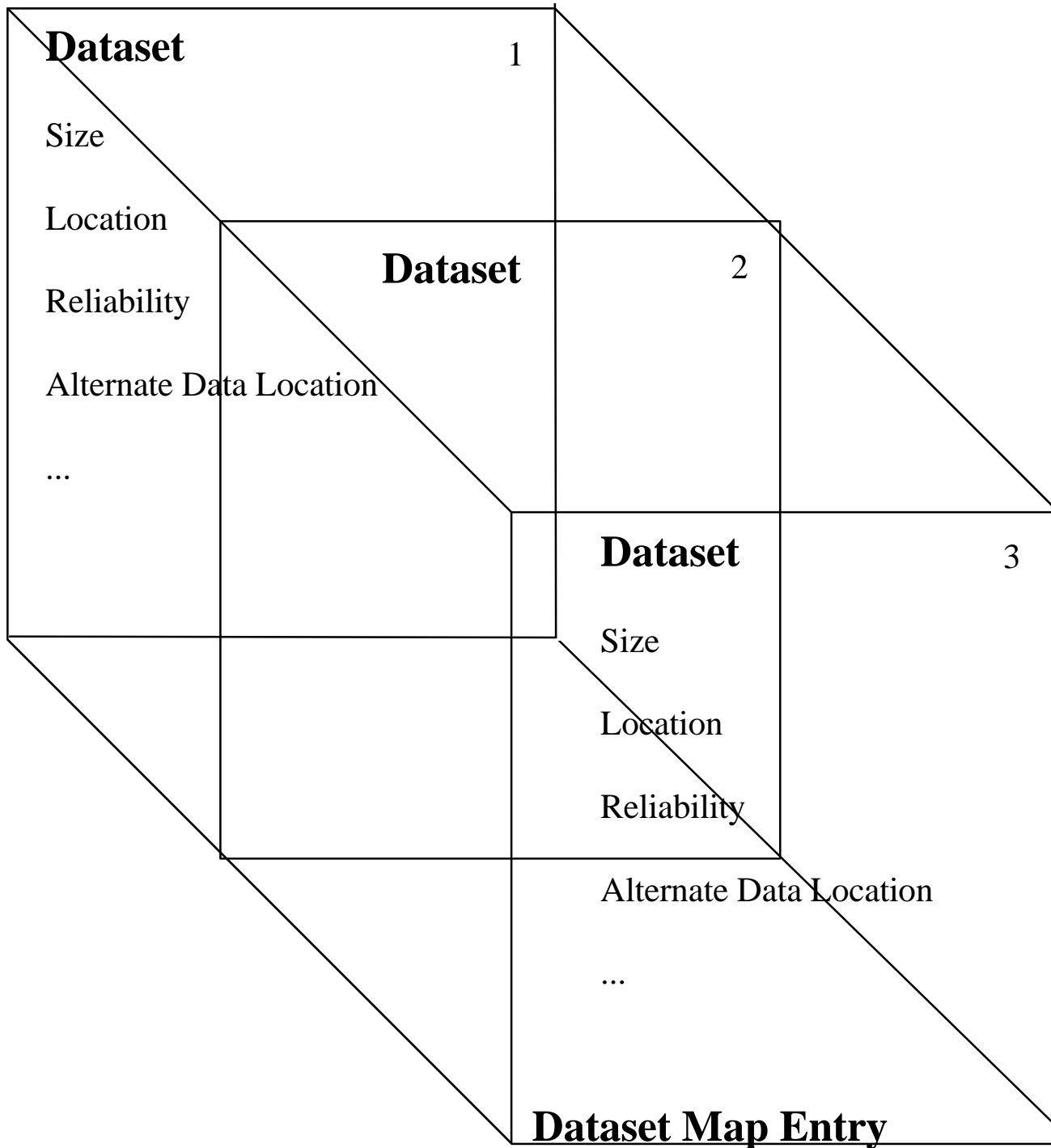Over time, can be trained to maintain a set of "useful data"
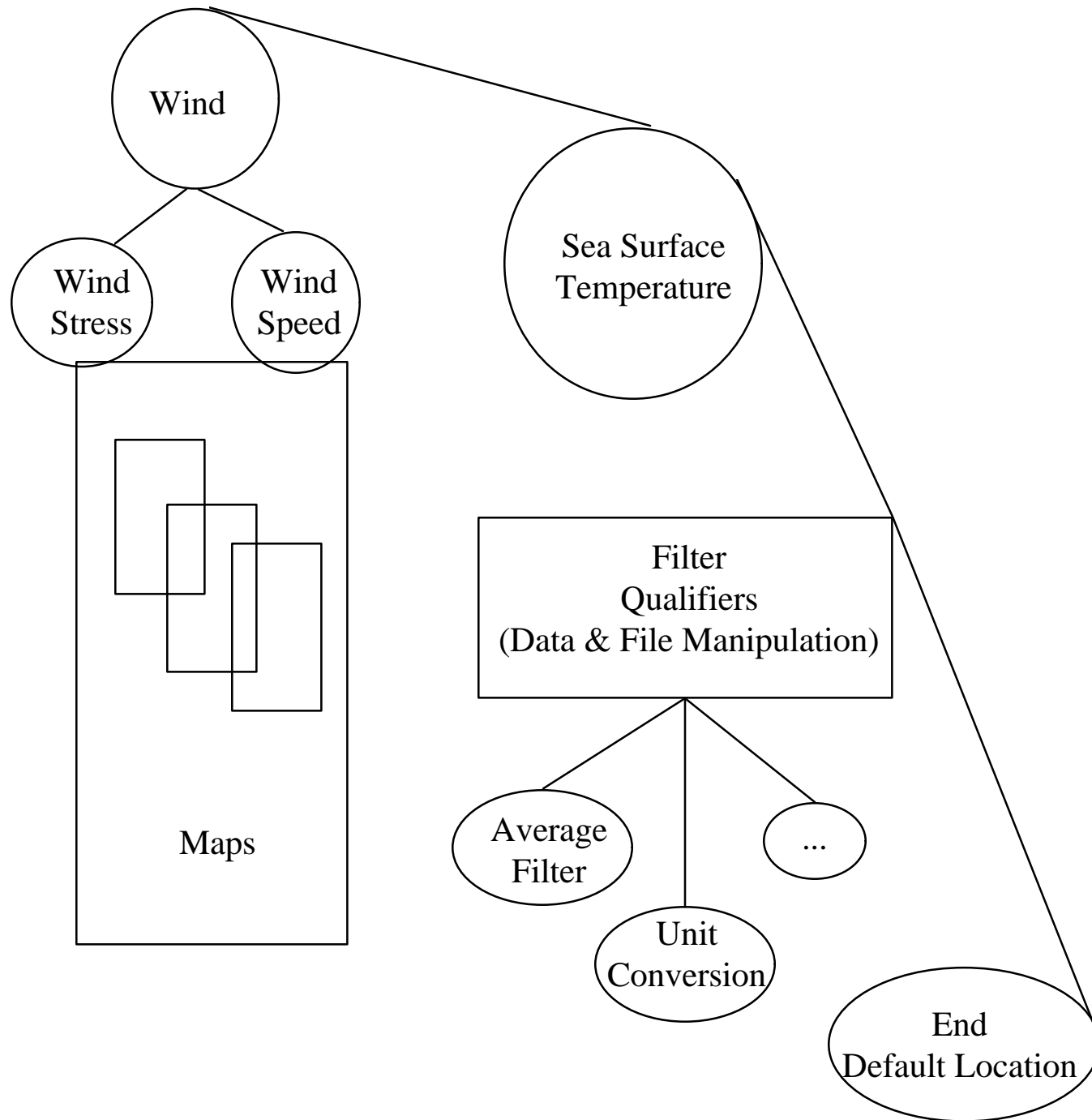
Data sets can be arbitrarily organized

VDADCs can be organized in a hierarchy

The search capability can be viewed as an optional product

Front End V0 Client

Data Picker

Browser

Data

DAAC

DAAC

DAAC

VDADC
Cache
File retrieval
Disk storage
CDs
Tapes

Front End Client

Data Picker

Browser

Workstation

Physical Data Center

DAAC
ADC
DADC
...

Physical Storage Media

Data

Data

Data

Virtual Data Center

VDADC

Data Pointers

Saved Data

VDADC Data Set

Dataset 1

Size

Location

Reliability

Alternate Data Location

...

Dataset 2

Dataset 3

Size

Location

Reliability

Alternate Data Location

...

Dataset Map Entry

Description

Locations

Valid dates

Valid times

Averages held

Records

Missing data

Statistical significance

Filter to convert to standard

Further Possibilities for Dataset Map Entry

III. Correlative Browse Studies on Global Data Sets

Ruixin Yang, Menas Kafatos

Center for Earth Observing and Space Research

Institute of Computational Sciences and Informatics

George Mason University

Data Source: ISLSCP Initiative I CD-ROM set by the GSFC DAAC

Data Structure:
- Spatial: One degree by one degree
- Temporal: Fixed; monthly 3(6)-hourly mean; instantaneous
- Format: ASCII or compressed ASCII
- Parameter: one data set for each parameter and each time

Data Sets Used:
- Sea Surface Temperature, monthly mean
- Sea Ice, monthly mean
- Cloud Product, monthly mean
- Mean Sea level Pressure, monthly 6-hourly
- Mean Wind at 10-M, monthly 6-hourly
- Mean Wind Stress, monthly 6-hourly

Tools:

- Linkwinds by JPL
- GrADS by COLA for wind vectors
- Splus, commercial statistical package.

Contacts

Prof. Menas Kafatos
703-993-1997
mkafatos@compton.gmu.edu

Duane King
703-993-1990
king@gmu.edu

Dr. Ruixin Yang
703-993-3615
yang@hubble.gmu.edu

Henry Wolf
IRM Associates
703-448-3394
hwolf@vms1.gmu.edu